**Proposal for Data Science Project**

**Deliverable 1**: 04/11/21

**Project Title**: Predicting potential employees’ performance rating before hiring.

**Data Retrieved From**: Kaggle: IBM HR Analytics Employee Attrition & Performance: <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

**Potential Data Fields**:

**Age** – Employee age

**Attrition** – Whether the employee is still in position

**Business Travel** -The frequency of business travel

**Daily Rate** - An internal measure in cost of employment per day (e.g., salary, insurance, logistics, overhead, etc.)

**Department** - Different departments in IBM

**Distance From Home** - The distance from employee’s home to company

**Education** - The education level of the employees (1: Below College, 2: College, 3: Bachelor 4: Master 5: Doctor)

**Education Field** - The focus major of the employees

**Employee Number** - Employee ID

**Environment Satisfaction** - The level of satisfaction that employee have on their working environment. (1: 'Low'; 2: 'Medium'; 3: 'High'; 4: 'Very High')

**Gender** - Female or Male

**Hourly Rate** - An internal measure in cost of employment per hour (e.g., salary, insurance, logistics, overhead, etc.)

**Job Involvement** - a state of psychological identification with work—or the degree to which a job is central to a person's identity. (1: 'Low'; 2: 'Medium'; 3: 'High'; 4: 'Very High')

**Job Level** - set the responsibility level and expectations of roles.

**Job Role** - The different functions you fill within the organization.

**Job Satisfaction** - How satisfied an employee. (1: Low, 2: Medium, 3: High, 4: Very High)

**Marital Status** - Employee Marital Status

**Monthly Income** - Monthly Salary

**Monthly Rate** - An internal measure in cost of employment per month (e.g., salary, insurance, logistics, overhead, etc.)

**Num Companies Worked** - number of previous employers.

**Over Time** - length of extended working hours

**Percent Salary Hike** - The parentage of change in salary between 2 year (2017, 2018).

**Performance Rating** - Numerical Value - ERFORMANCE RATING (1: Low, 2: Medium, 3: High, 4: Very High)

**Relationship Satisfaction** - Numerical Value – How satisfied an employee about the company (1: Low, 2: Medium, 3: High, 4: Very High)

**Standard Hours** - How many hours

**Stock Option Level** - numerical value - stock options. The company stocks people own from this company.

**Total Working Years** - Total number of years at the company

**Training Times Last Year** - numerical value - hours spent training

**Work Life Balance** - numerical value - time spent between work and outside (1: bad;  
2: good; 3: better; 4: best)

**Years at Company** - numerical value - total number of years at the company

**Years in Current Role** - years in current role

**Years Since Last Promotion** - numerical value - last promotion

**Years with Curr Manager** - numerical value - years spent with current manager.

**Business Problem Addressed**:

Employees’ job performance rating is, arguably, one of the most effective metrics that evaluates employee’s contributions. In most cases, to obtain an employee’s performance rating would require a continues assessment of their work. However, in the light of the situation where hiring new employees bares high opportunity cost. Companies could benefit from predicting one’s performance rating before committing to hiring. Other than that, in finding more suitable employees, companies could rely on indicators that could, in some degrees, accurately reflect one’s aptness. Therefore, to reduce risk and uncertainty in hiring, the company needs a classification model that could predict one’s performance with information before hiring.

**Use Scenario of Result**: We will analyze how the education field, distance from home job involvement, job role, job satisfaction, work life balance, etc. impact the performance rating. This helps us to understand the employee's behaviors in order to predict their future performance. Thereby, this will improve the hiring process as well as decrease uncertainties in human resource management, operation, and administration.

**Data Instance and Useful Features**: Since we are trying to predict the working performance right before hiring, we will mainly focus on pre-work variables. The main data variables we will include in our instance are Education, Distance from home, Education Field and Monthly Income, etc.

**Target Variable**: Performance Rating - Numerical Value - ERFORMANCE RATING (1: Low, 2: Medium, 3: High, 4: Very High)

**Added Business Value**: The company would gain the ability to access one’s potential performance rating before committing to hire that person. Hence, to some degrees of extend, the company mitigated risks in hiring incompatible people.